From General to Clinical: Adapting Foundation Models for Medical Images (Winter semester 25/26)

Abstract

Deep learning models from CNNs [6] to Vision Transformers [4, 10], have achieved impressive results for a variety of benchmarks and applications. However, real-world deployment, especially in medical imaging, remains difficult due to unseen variations in scanners, patient demographics, and pathological conditions. To address this issue, recent works focus on foundation models: large-scale models pretrained on diverse data using self- or weakly supervised learning [2]. CLIP-based [12] models learn general-purpose representations that transfer well to downstream tasks. For the context of medical imaging, specific fine-tuned models such as MedSAM [11] and Biomed-CLIP [14] demonstrate strong potential. However, adapting foundation models to new domains or tasks often remains challenging. Although domain generalization and adaptation approaches exist, they may overfit or require target data. A recent paradigm, Test-time adaptation [1, 13] mitigates this by adapting alongside inference to the unseen data. The techniques include adapting dynamically through model updates or prompt modifications that support more flexible deployment.

This seminar will explore the foundations and capabilities of foundation models, with a focus on pretraining strategies, representational generalization, and efficient adaptation methods. We will examine state-of-the-art foundation models and discuss how they can be adapted to downstream tasks, especially in medical imaging, through techniques such as self-supervised learning, cross-domain transfer, and parameter-efficient tuning.

Kick-off Zoom Link: July 10th, 3pm Meeting ID: 401 232 4259 Passcode: 389389
TUM Campus Number: 0000001178

Objective

We will review state-of-the-art foundation models and their practical adaptations for medical imaging to cover: **(i)** *What to adapt:* identify which model parameters to adapt, explore common strategies for adaptation **(ii)** *How to adapt:* The various strategies by which the models can be adapted to the target data **(iii)** *Evaluation for clinical context:* discuss methods for handling distributional shifts that mirror real-world clinical scenarios.

Teaching and learning methods

We will begin by introducing foundation models, tracing their evolution from natural language processing and computer vision to medical data analysis. Next, we will review state-of-the-art foundation models [3, 7, 8, 12] and adaptation methods such as [5, 9, 13]. Students will select a research paper, independently work on it, and make an oral presentation. In addition, students will discuss and present applications of the Foundation models in the context of medical imaging in a final poster session. Throughout the seminar, we will have discussions about the challenges, limitations, and future directions of adapting foundation models.

Contact. Sameer Ambekar or Dr. Laura Daza

References

[1] S. Ambekar and et al. Probabilistic test-time generalization by variational neighbor-labeling. *arXiv*, 2023.

[2] R. Bommasani and et al. On the opportunities and risks of foundation models. *arXiv*, 2021.

[3] M. Caron and et al. Emerging properties in self-supervised vision transformers. *ICCV*, 2021.

[4] A. Dosovitskiy and et al. An image is worth 16×16 words: Transformers for image recognition at scale. *ICLR*, 2021.

[5] C. Ge and et al. Domain adaptation via prompt learning. *IEEE TNNLS*, 2023.

[6] K. He and et al. Deep residual learning for image recognition. In *CVPR*, 2016.

[7] K. He and et al. Masked autoencoders are scalable vision learners. *CVPR*, 2022.

[8] A. Kirillov and et al. Segment anything. *ICCV*, 2023.

[9] Y. Liu and et al. Ttt++: When does self-supervised test-time training fail or thrive? In *NeurIPS*, 2021.

[10] Z. Liu and et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, 2021.

[11] J. Ma and et al. Medsam: Segment anything model for medical images. *arXiv*, 2023.

[12] A. Radford and et al. Learning transferable visual models from natural language supervision. *ICML*, 2021.

[13] D. Wang and et al. Tent: Fully test-time adaptation by entropy minimization. In *ICLR*, 2021.

[14] Y. Zhang and et al. Biomedclip: Open biomedical contrastive language-image pretraining. *arXiv*, 2023.